Downloaded from orbit.dtu.dk on: Apr 19, 2024

DTU Library

Stochastic Economic Dispatch with Wind using Versatile Probability Distribution and L-BFGS-B Based Dual Decomposition

Huang, Shaojun; Sun, Yuanzhang; Wu, Qiuwei

Published in:
IEEE Transactions on Power Systems

Link to article, DOI:
10.1109/TPWRS.2018.2834433

Publication date:
2018

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Huang, S., Sun, Y., & Wu, Q. (2018). Stochastic Economic Dispatch with Wind using Versatile Probability Distribution and L-BFGS-B Based Dual Decomposition. *IEEE Transactions on Power Systems*, *33*(6), 6254 - 6263. https://doi.org/10.1109/TPWRS.2018.2834433

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Stochastic Economic Dispatch with Wind using Versatile Probability Distribution and L-BFGS-B Based Dual Decomposition

Shaojun Huang, *Member, IEEE*, Yuanzhang Sun, *Senior Member, IEEE, and* Qiuwei Wu, *Senior Member, IEEE*

***Abstract*—This paper focuses on economic dispatch (ED) in power systems with intermittent wind power, which is a very critical issue in future power systems. A stochastic ED problem is formed based on the recently proposed versatile probability distribution (VPD) of wind power. The problem is then analyzed and proved to be strictly convex. Although such convex optimization is tractable in many cases, it may take a long time to solve due to its large scale. This paper proposes a dual decomposition method to decompose the large problem. Then two methods are employed to solve the decomposed problem, namely, the subgradient method and a faster method, limited-memory BFGS with box constraints (L-BFGS-B, a quasi-Newton method). Case studies were conducted to verify the efficiency of the dual decomposition and L-BFGS-B method for solving the stochastic ED problem.**

***Index Terms*— dual decomposition, economic dispatch (ED), limited-memory BFGS with box constraints (L-BFGS-B), stochastic program, sub-gradient method, versatile probability distribution, wind power.**

I. Nomenclature

Sets

N_l	set of lines
N_g	set of conventional generators
N_T	set of planning periods
N_w	set of wind farms

Parameters

$G_{k,t}^{g}$	power transfer distribution factor (PTDF) regarding generators
$G_{k,t}^{w}, G_{k,t}^{d}$	PTDF regarding wind farms/loads
R_t^d, R_t^u	total down/up reserve requirement
a, b, c	generation cost coefficients
c^-, c^+	confidence level
$d_{i,t}$	loads (kW)
k^{ov}, k^{un}	over/under estimate cost of wind farms
$l_{k,t}^{\max}$	loading limit of lines
m	L-BFGS-B: number of stored historical points
$p_i^{g\min}, p_i^{g\max}$	generator min/max output power
$p_i^{w\min}, p_i^{w\max}$	wind farm min/max output power
$r_i^{d\max}, r_i^{u\max}$	maximum down/up ramping limit of generators
α, β, γ	VPD parameters
α	step size (subgradient and L-BFGS-B method)
ε	a small penalty coefficient
θ	parameter of L-BFGS-B method

Variables

$p_{i,t}^{g}$	scheduled power of normal generators(kW)
$p_{i,t}^{w}$	scheduled power of wind farms (kW)
$p_{i,t}^{wa}$	(random) actual power of wind farms (kW)
$r_{i,t}^{d}$	auxiliary variable: down reserve of generators
$r_{i,t}^{u}$	auxiliary variable: up reserve of generators
λ_t	dual variable of load balance constraint
μ_t^-, μ_t^+	dual variable of wind power chance constraint
v_t^-, v_t^+	dual variable of total reserve limit
$\omega_{k,t}^-, \omega_{k,t}^+$	dual variable of line loading limit

Other Symbols

$\Pr\{*\}$	probability of *

II. Introduction

WIND power has been significantly deployed in power systems of many countries in recent years in order to reduce CO_2 emissions. The large amount of intermittent renewable energy, including wind power and solar power, challenges the power systems through the planning stage to the real time operation stage from many aspects. This paper focuses on the economic dispatch (ED) problem [1], [2] in power systems with high penetration of wind power.

The ED is an optimal power flow (OPF) problem with the objective to minimize the total generation cost subject to security constraints of the power system and the limits of the generators. The traditional ED problems without renewable energy are usually deterministic optimization problems [3]. They can also be stochastic programming problems if the loads are considered as random variables [4]. Being non-dispatchable, wind power scheduling is dependent on the forecast and consequently has stochastic features. The ED

S. Huang and Q. Wu are with the Center for Electric Power and Energy (CEE), Department of Electrical Engineering, Technical University of Denmark (DTU), 2800, Kgs. Lyngby, Denmark (e-mail: shuang@elektro.dtu.dk; qw@elektro.dtu.dk).

Y. Sun is with the School of Electrical Engineering, Wuhan University, Wuhan, 430077, China (e-mail: yzsun@whu.edu.cn).

problem with wind power has been frequently modeled as a stochastic programing problem in previous studies [5-10]. In [5], the Weibull probability distribution function (PDF) was employed to model the random wind power. Chance constraints of the wind power were considered in the ED problem in [6] and they were transformed into deterministic constraints by the Gaussian distribution. However, due to the fact that wind power is bounded between zero and the rated power, and the forecast accuracy is largely dependent on the forecast time scale, two statistical indexes, i.e., skewness and kurtosis, are very important in correctly modeling the random wind power [7]. Distributions that have shape parameters reflecting the skewness and kurtosis include the Beta distribution and Cauchy distribution [7]. The histogram method [8] and scenario-based method [9], [10] can be employed for modeling the wind power in the ED problem. However, there is a tradeoff between choosing a large number of bins or scenarios to improve the accuracy and a small number of those to reduce the complexity of the problem.

In [11], [12] a three-parameter distribution, named as versatile probability distribution (VPD), was proposed, which has a few nice features: (1) it has three parameters and therefore can better capture the skewness and kurtosis of the random wind power compared to the normal, Beta and Cauchy distributions; (2) it has analytical forms for the PDF, cumulative distribution function (CDF) and reverse CDF. The second feature is further employed in this paper to analyze the convexity, first- and seconder-order derivatives of the objective function and speed up the solution process.

As the stochastic ED problem is usually large scale due to a large amount of generators, buses, planning periods and nonlinear terms containing stochastic variables, the direct method, such as sequential linear programming (SLP) [12] or interior point algorithm, may be very slow in searching the optimal point. Heuristic algorithms, such as particle swarm optimization [13]–[15], are also employed to solve such problems. This paper will employ the dual decomposition method [3], [16] to decompose the large and difficult problem into many small sub-problems, each of which handles one generator (wind farm) scheduling. Furthermore, the dual variables (Lagrange multipliers) in the master problem can be updated using the subgradient method [17] or quasi-Newton method, such as Broyden–Fletcher–Goldfarb–Shanno (BFGS) [3], [18]–[20]. This paper will investigate another quasi-Newton method, namely, the limited-memory BFGS with box constraints (L-BFGS-B) method [21], which can save memory for such large scale problems and improve computation efficiency.

The main contributions of this paper are summarized as follows: (1) Prove that the VPD based stochastic ED with linear power flow constraints is a strictly convex optimization. Furthermore, the first- and second-order derivatives of the objective function are derived which can speed up the interior point algorithm for solving the subproblems of the dual decomposition method. These results provide more fundamental insights of the VPD based optimization problems compared to previous studies [11], [12], and they are critical for nonlinear problems being tractable. (2) Propose the dual decomposition method for decomposing the stochastic ED problem with wind. The dual decomposition method is very important here because the stochastic ED problem with wind is a very large and complicate nonlinear problem. Although the BFGS based dual decomposition method has already been proposed in the previous study [3], the main problem solved in [3] is a deterministic quadratic problem. However, the stochastic ED problem with wind is highly nonlinear and very difficult to solve directly even though it is a convex optimization as proved in this paper. Further explanation can be seen in IV.A. (3) Further solve the decomposed problem by the subgradient method (proposed in previous work, as the basic case) and the proposed L-BFGS-B method, respectively; discuss and compare these two methods. There are two important features of the L-BFGS-B method compared to the BFGS method used in [3]. One is that the L-BFGS-B method can save memory usage and have fast computation because it only saves and computes a limited number of n-vectors instead of $n \times n$ full matrix as in the BFGS method (n is the number of Lagrange multipliers and can be as high as a few thousands). The other one is that the L-BFGS-B method can handle box constraints. For instance, the Lagrange multipliers of the inequality constraints of the stochastic ED problem are required to be positive and the L-BFGS-B method can well handle this requirement – even can take advantage of this requirement to reduce the computation.

The rest of the paper is organized as follows. The mathematical formulation of the VPD and the modeling of the stochastic ED problem are presented in Section III. The convexity analysis of the formulated stochastic ED problem and its solution using the dual decomposition and subgradient method are presented in Section IV. Then, the L-BFGS-B method, which is fast and memory-efficient, for solving the same problem is introduced in Section V. In Section VI, case studies are presented and discussed. The paper ends with the conclusions.

III. Modeling of Stochastic Wind Power Production and Its Application for Economic Dispatch

In this section, the mathematical formulation of the VPD for stochastic wind power production is presented first. Afterwards, the modeling of the ED problem using the VPD is presented.

A. Modeling Wind Power Production with Versatile Probabilistic Distribution

In [11], [12], the VPD was introduced in order to model stochastic wind power production more accurately than the traditional distribution functions, such as the normal distribution (Gaussian) and beta distribution. The VPD has three parameters, which make it more flexible and powerful in terms of modeling stochastic wind power production.

The PDF of the VPD can be expressed as,

$$f(x) = \frac{\alpha\beta e^{-\alpha(x-\gamma)}}{(1+e^{-\alpha(x-\gamma)})^{\beta+1}} \ , \tag{1}$$

where α , β and γ are parameters, and $\alpha > 0$, $\beta > 0$. Its CDF is,

$$F(x) = (1 + e^{-\alpha(x-\gamma)})^{-\beta} \quad . \tag{2}$$

It can be verified that $f(x) > 0$ (when $x \to \pm\infty$, $f \to 0$), and the derivative of F is f . The inverse F can be found as,

$$F^{-1}(c) = \begin{cases} -\infty & c = 0 \\ \gamma - \dfrac{1}{\alpha}\ln(c^{-1/\beta} - 1) & 0 < c < 1 \\ +\infty & c = 1 \end{cases} . \tag{3}$$

According to (1), α determines the kurtosis of the PDF, while β determines the skewness, and γ determines the displacement of the function. It can be proven that, when β =1, the function is symmetric about $x = \gamma$, and the function has peak value $\alpha / 4$ at $x = \gamma$; however, when $\beta \neq 1$, the peak value is not at $x = \gamma$.

The analytical forms (consisting of elementary functions only) of the PDF, CDF and inverse CDF give a lot of advantages in modeling ED problems as well as solving the ED problems. On the contrary, the traditional distributions employed to represent the stochastic wind power production, such as the normal distribution or beta distribution, do not have analytical forms for the CDF and its inverse. In this paper, the advantages of the VPD will be explored in modeling and solving the ED problems.

As pointed out by [22], the forecast error (forecast value minus the actual one) of the wind power production might have different distributions, i.e., different α , β and γ in case of the VPD, for different expected wind power production level in the range 0~1 p.u. (1 p.u. means the rated power). The reason is that the actual wind power production should be bounded in 0~1 p.u.. Hence, for the ED problem with multiple planning periods, the stochastic wind production of different periods will be represented by different VPDs. Fig. 1 shows three VPDs for three different expected value (close to 0, 0.5 and 1 p.u., respectively). Specific α , β and γ values can be seen in Section VI.A (Table IV).

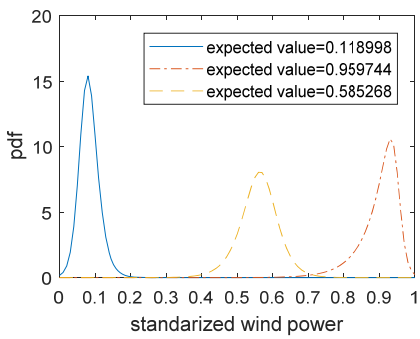

Fig. 1. Illustration of VPDs for three different forecast values of the wind power

B. Stochastic ED Model with Wind Power

An ED problem is to meet load demands by scheduling generators with the minimum cost and respecting network security constraints at the same time. When the stochastic wind power production is included in the ED problem, the cost function and constraints should be modified compared to the traditional ED problem. An ED problem with stochastic wind power production can be modeled as follows.

$$\min \sum_{t \in N_T} \{ \sum_{i \in N_g} C_i^g(p_{i,t}^g, r_{i,t}^d, r_{i,t}^u) + \sum_{i \in N_w} C_{i,t}^w(p_{i,t}^w) \} \tag{4}$$

s.t.

$$\sum_{i \in N_g} p_{i,t}^g + \sum_{i \in N_w} p_{i,t}^w - \sum_{i \in N_d} d_{i,t} = 0, \forall t \in N_T, (\lambda_t) \tag{5}$$

$$\sum_{i \in N_g} r_{i,t}^d \geq R_t^d, \forall t \in N_T, (v_t^-) \tag{6}$$

$$\sum_{i \in N_g} r_{i,t}^u \geq R_t^u, \forall t \in N_T, (v_t^+) \tag{7}$$

$$\begin{gathered} -l_{k,t}^{\max} \leq \sum_{i \in N_g} G_{k,t}^g p_{i,t}^g + \sum_{i \in N_w} G_{k,t}^w p_{i,t}^w - \sum_{i \in N_d} G_{k,t}^d d_{i,t} \leq l_{k,t}^{\max}, \\ \forall k \in N_l, t \in N_T, (\omega_{k,t}^-, \omega_{k,t}^+) \end{gathered} \tag{8}$$

$$\Pr\left\{ \sum_{i \in N_w} (p_{i,t}^{wa} - p_{i,t}^w) \leq \sum_{i \in N_g} r_{i,t}^d \right\} \geq c^-, \forall t \in N_T, (\mu_t^-) \tag{9}$$

$$\Pr\left\{ \sum_{i \in N_w} (p_{i,t}^w - p_{i,t}^{wa}) \leq \sum_{i \in N_g} r_{i,t}^u \right\} \geq c^+, \forall t \in N_T, (\mu_t^+) \tag{10}$$

$$r_{i,t}^d \leq p_{i,t}^g - p_i^{g\min}, \forall i \in N_g, t \in N_T \tag{11}$$

$$r_{i,t}^d \leq r_i^{d\max}, \forall i \in N_g, t \in N_T \tag{12}$$

$$r_{i,t}^u \leq p_i^{g\max} - p_{i,t}^g, \forall i \in N_g, t \in N_T \tag{13}$$

$$r_{i,t}^u \leq r_i^{u\max}, \forall i \in N_g, t \in N_T \tag{14}$$

$$p_i^{g\min} \leq p_{i,t}^g \leq p_i^{g\max}, \forall i \in N_g, t \in N_T \tag{15}$$

$$-r_{i,t}^{d\max} \leq p_{i,t}^g - p_{i,t-1}^g \leq r_{i,t}^{u\max}, \forall i \in N_g, t \in N_T \setminus \{1\} \tag{16}$$

$$p_i^{w\min} \leq p_{i,t}^w \leq p_i^{w\max}, \forall i \in N_w, t \in N_T \tag{17}$$

In the above model, $p_{i,t}^{wa}$ is the actual wind power production (curtailment is not considered in this paper), which is stochastic and has a VPD. Therefore, the cost function $C_{i,t}^w(p_{i,t}^w)$ is determined by the distribution of $p_{i,t}^{wa}$, and (9) and (10) are chance constraints. Therefore, the ED model is a stochastic programming problem, which is usually hard to solve. Fortunately, the VPD has analytical forms for the PDF, CDF and inverse CDF, which can help transform the above stochastic programming problem into a deterministic programming problem (Section III.C).

The cost function (4) consists of two parts, namely the cost associated with conventional generators and the one associated with wind farms. Constraint (5) forces a load balance, constraints (6) and (7) ensure the total downward and upward generation reserves respectively, and constraint (8) concerns the line loading limits. Constraints (9) and (10) make sure that the downward and upward reserves can, with a certain confidence level (c^- and c^+), cover the needs of the wind farms

due to the mismatch of scheduling and the real dispatches. Constraints (11) and (12) are related to the downward reserve, while (13) and (14) are related to the upward reserve. Constraints (15) and (16) are the power limits and ramping limits of the conventional power plants. Finally, (17) gives the power limits of the wind farms.

C. Equivalent Deterministic Model

The cost function of conventional power plants in (4), i.e., $C_i^g(p_{i,t}^g, r_{i,t}^d, r_{i,t}^u)$, can be represented by a quadratic function as,

$$C_i^g(p_{i,t}^g, r_{i,t}^d, r_{i,t}^u) = a_i p_{i,t}^{g\,2} + b_i p_{i,t}^g + c_i + \varepsilon(r_{i,t}^{d\,2} + r_{i,t}^{u\,2}) \ . \quad (18)$$

The term associated with ε (a small positive value) is a penalty function to maintain strict convexity of the cost function, which is very important in the convergence of the dual decomposition method (see Section IV.B and V). Usually, $a, b, c, \varepsilon > 0$.

The cost function of wind farms, i.e., $C_{i,t}^w(p_{i,t}^w)$, can be represented by,

$$\begin{aligned} C_{i,t}^w(p_{i,t}^w) = {} & k^{ov}\int_0^{p_{i,t}^w}(p_{i,t}^w - p_{i,t}^{wa}) f_{i,t}(p_{i,t}^{wa}) dp_{i,t}^{wa} \\ & + k^{ud}\int_{p_{i,t}^w}^{p_i^{w\max}}(p_{i,t}^{wa} - p_{i,t}^w) f_{i,t}(p_{i,t}^{wa}) dp_{i,t}^{wa} \ . \quad (19) \\ & + \varepsilon(p_{i,t}^w - p_{i,t}^{we})^2 \end{aligned}$$

It can be seen from (19) that, if the distribution function $f_{i,t}$ is given, the cost function is determined. A small ε can keep the strict convexity of the cost function even when $p_{i,t}^w$ is far away from $p_{i,t}^{we}$, i.e., the expected value of $p_{i,t}^{wa}$ (see Section IV.A).

All constraints except (9) and (10) are linear. According to [12], (9) and (10) can be transformed into linear constraints as well, as the cumulative distribution of $\sum_{i\in N} p_{i,t}^{wa}$, i.e., $F_{\Sigma,t}$, is assumed to be known. Accordingly, (9) and (10) are equivalent to,

$$\sum_{i\in N_w} p_{i,t}^w + \sum_{i\in N_g} r_{i,t}^d \le F_{\Sigma,t}^{-1}(c^-), \forall t \in N_T, (\mu_t^-) \ , \quad (20)$$

$$\sum_{i\in N_w} p_{i,t}^w - \sum_{i\in N_g} r_{i,t}^u \le F_{\Sigma,t}^{-1}(1-c^+), \forall t \in N_T, (\mu_t^+) \ . \quad (21)$$

To sum up, the equivalent deterministic ED problem is: (4) s.t. (5)-(8), (11)-(17) and (20)-(21). The cost function still contains nonelementary functions (integral function); however, all constraints are affine functions. In the next two sections, the problem will be analyzed and the algorithm to solve it will be provided.

IV. Convexity Analysis and Dual Decomposition

A. Convexity of the ED problem

As pointed out in Section III.C, only the cost function needs to be analyzed, as all constraints are linear. In fact, the cost function is totally separable regarding different generators, wind farms and time periods. Therefore, a convexity analysis of individual cost functions (associated with one variable) is sufficient. It is obvious that the cost function associated with the conventional generators (variable $p_{i,t}^g, r_{i,t}^d, r_{i,t}^u$ respectively) is quadratic. With $a, \varepsilon > 0$, it is strictly convex. The focus will be given to $C_{i,t}^w(p_{i,t}^w)$.

For given $i \in N_w, t \in N_T$, the first and second order derivatives of $C_{i,t}^w(p_{i,t}^w)$ are,

$$\begin{aligned} \frac{dC_{i,t}^w}{dp_{i,t}^w} = {} & k^{ud}(F_{i,t}(p_{i,t}^w) - F_{i,t}(p_{i,t}^{w\max})) \\ & + k^{ov}(F_{i,t}(p_{i,t}^w) - F_{i,t}(0)) \quad , \quad (22) \\ & + 2\varepsilon(p_{i,t}^w - p_{i,t}^{we}) \end{aligned}$$

$$\frac{d^2C_{i,t}^w}{dp_{i,t}^{w\,2}} = (k^{ud} + k^{ov}) f_{i,t}(p_{i,t}^w) + 2\varepsilon \ . \quad (23)$$

It can be seen that the second order derivative is always positive as the density function $f_{i,t}$ is always positive and $k^{ov}, k^{ud}, \varepsilon > 0$. Hence, $C_{i,t}^w(p_{i,t}^w)$ is strictly convex. In addition, when $p_{i,t}^w$ is close to $p_{i,t}^{we}$, $f_{i,t}$ is significant and dominates in the second order derivative; however, when it is far away from $p_{i,t}^{we}$, $f_{i,t}$ varnishes and ε dominates in the second order derivative. This is why ε is needed for keeping the strict convexity. Although the primary function is non-elementary, its first and second order derivatives are elementary functions.

From the above analysis, it can be concluded that the deterministic ED problem is a strict convex optimization, although its cost function still contains non-elementary functions. This problem can be solved by, e.g., the interior method [23], [24], and one can provide the first and second order derivatives to speed up the solving process.

However, when there are a large number of generators and wind farms, this convex problem can still be hard to solve. Because its objective function has non-elementary functions (integrals), efficient commercial solvers, such as CVX [25], are not applicable. The Matlab solver "fmincon" supports non-elementary functions; however, if the entire problem is fed into the solver, it will be very slow. The idea in this paper is to decompose the problem and solve each small problem by "fmincon" with the interior point algorithm.

B. Dual Decomposition

Only (5)-(8) and (20)-(21) of the ED problem are coupling variables of different generators and wind farms, while the cost function and other constraints are totally separable. It is suitable to use the dual decomposition method [16]. Only the coupling constraints need to be dualized. As the primal problem is strictly convex (proven in Section IV.A), it is equivalent to its dual problem.

The dual problem can be expressed through the "max min" formulation as,

$$\max_{\lambda_t \text{ and } (v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+) \geq 0} \quad \min_{p_{i,t}^g, r_{i,t}^d, r_{i,t}^u, p_{i,t}^w}$$

$$L(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+, p_{i,t}^g, p_{i,t}^w, r_{i,t}^d, r_{i,t}^u)$$

$$= \sum_{t \in N_T} \{ \sum_{i \in N_g} C_i^g(p_{i,t}^g, r_{i,t}^d, r_{i,t}^u) + \sum_{i \in N_w} C_{i,t}^w(p_{i,t}^w)$$

$$+ \lambda_t (\sum_{i \in N_g} p_{i,t}^g + \sum_{i \in N_w} p_{i,t}^w - \sum_{i \in N_d} d_{i,t})$$

$$+ v_t^-(-\sum_{i \in N_g} r_{i,t}^d + R_t^d) + v_t^+(-\sum_{i \in N_g} r_{i,t}^u + R_t^u)$$

$$+ \sum_k \omega_{k,t}^-(-I_k^{\max} - \sum_{i \in N_g} G_{k,t}^g p_{i,t}^g - \sum_{i \in N_w} G_{k,t}^w p_{i,t}^w + \sum_{i \in N_d} G_{k,t}^d d_{i,t})$$

$$+ \sum_k \omega_{k,t}^+(-I_k^{\max} + \sum_{i \in N_g} G_{k,t}^g p_{i,t}^g + \sum_{i \in N_w} G_{k,t}^w p_{i,t}^w - \sum_{i \in N_d} G_{k,t}^d d_{i,t})$$

$$+ \mu_t^-(\sum_{i \in N_w} p_{i,t}^w + \sum_{i \in N_g} r_{i,t}^d - F_{\Sigma,t}^{-1}(c^-))$$

$$+ \mu_t^+(\sum_{i \in N_w} p_{i,t}^w - \sum_{i \in N_g} r_{i,t}^u - F_{\Sigma,t}^{-1}(1-c^+))\} \quad , (24)$$

s.t. (11)-(17).

The dual problem is a two-layer problem: the outer layer, viewed as the master problem, is the maximization part; the inner layer, viewed as the sub problem, is the minimization part. It can be seen that the sub problem can be separated into small problems regarding each generator and wind farm ($\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$ are fixed for the sub problem). As the sub problem is separable and each small problem, which is strictly convex, is relatively easy to solve (can use the interior point method with given analytical first and second order derivatives), the difficulty is now on solving the master problem. In the following sections, the master problem is to be solved by the subgradient method and quasi Newton method.

C. Subgradient Method for the Master Problem

The master problem in Subsection IV.B is to maximize the objective function within domain $v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+ \geq 0$. The master problem can be expressed as,

$$\max_{\lambda_t \text{ and } (v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+) \geq 0} M(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+) \quad .(25)$$

Function $M()$ is evaluated by solving the sub problem for given $\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$ and it is equal to the optimal value of the sub problem, which means,

$$M(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+) = \min_{p_{i,t}^g, r_{i,t}^d, r_{i,t}^u, p_{i,t}^w} L(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+, p_{i,t}^g, p_{i,t}^w, r_{i,t}^d, r_{i,t}^u) \quad ,(26)$$

s.t. (11)-(17).

Because the sub-problem is strictly convex, $M()$ always has a single value for any given point, which means that $-M()$ is well defined in its domain. It can be proved that the function is continuous. Although $M()$ is not always differentiable in its domain, it has subgradient at every point, which makes it possible to use the subgradient method [17] to search the optimum point. The subgradients with respect to the variables, $\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$, are the mismatches of the corresponding constraints ((5)-(10)), respectively. The subgradient method can start with an initial point, e.g., ($\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$)=0, then updated by the subgradients multiplying a small step size α, till reach the optimal point $(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+)^*$. Details of the method are omitted for brevity.

It is interesting to know that, because the sub-problem is strictly convex, the subgradients of $M()$ are actually the gradients in most cases. Specifically, when the inequality constraints of the sub-problems are non-degenerate ("degenerate" means that the constraint is active and the Lagrange multiplier is zero at the optimum point) for a given point ($\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$), the subgradients of $M()$ are the gradients at that point. Therefore, in this case, the subgradient method is very close to the gradient method (the descent method).

V. Quasi-Newton Method for The Master Problem

Following the subgradient method introduced in Subsection IV.C, this section introduces a quasi-Newton method, namely, the L-BFGS-B method [21], to solve the master problem (25). The subgradient method needs a large number of iteration steps in many cases before reaching the optimal point, especially when there are many congestion points of the security constrained OPF (many nonzero Lagrange multipliers). Therefore, a fast method for searching the optimal point is needed in many cases.

As mentioned in Subsection IV.C, the subgradients of the objective function $M()$ are actually the gradients in most cases. Therefore, it is possible to estimate the Hessian matrix of $M()$ and using the Newton method to solve the master problem. However, it is time-costly to calculate an accurate Hessian matrix by numerical methods because one evaluation of $M()$ means solving the sub problem once. The L-BFGS-B method is suitable in this case, because it can use the information in the current iteration step (one evaluation of $M()$) and a few historical iteration steps to estimate an approximate Hessian matrix. It is also memory-efficient because it only stores information of a few steps, e.g., *m* steps, instead of a full Hessian matrix as in the Newton method and the BFGS method.

For convenience, the master problem is reformed as a minimization problem,

$$\min_x \; u(x), \quad (27)$$

where,

$$u(x) = -M(\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+) \quad ,$$

$$x \in \{x \mid x = [\lambda^T, v^{-T}, v^{+T}, \omega^{-T}, \omega^{+T}, \mu^{-T}, \mu^{+T}]^T, (v^-, v^+, \omega^-, \omega^+, \mu^-, \mu^+) \geq 0\} \quad ,$$

and $\lambda, v^-, v^+, \omega^-, \omega^+, \mu^-, \mu^+$ are vector forms of $\lambda_t, v_t^-, v_t^+, \omega_{k,t}^-, \omega_{k,t}^+, \mu_t^-, \mu_t^+$, respectively.

The algorithm is adapted to avoid too many function evaluations.

The Modified L-BFGS-B Algorithm:

1) Initialize: $x^{(0)}$ (e.g., 0), θ (e.g., 1000) and m (e.g., 5);

2) Evaluate $u(x)$ at $x^{(0)}$ by solving the sub-problem, evaluate the gradients of $u(x)$ at $x^{(0)}$, i.e., $g^{(0)}$ (mismatches of the corresponding constraints, notice the sign);

Let $k = 0$ and $m_1 = 0$;

3) Project $g^{(k)}$ to the free-variable sub space, denoted as $\hat{g}^{(k)}$. Free variables are the variables without limits, e.g., λ, or the lower limits are not bounding for the corresponding direction determined by $-g^{(k)}$;

If $\left\|\hat{g}^{(k)}\right\| \le \tau$, stop; otherwise,

4) Calculate $B^{(k)}$ and $\hat{B}^{(k)}$:

$B^{(k)} = \theta I - WMW^T$, where,

$$W = \begin{bmatrix} Y & \theta S \end{bmatrix},\ M = \begin{bmatrix} -D & L^T \\ L & \theta S^T S \end{bmatrix},$$

$$Y = [y^{(k-m_1)}, y^{(k-m_1+1)}, y^{(k-m_1+2)}, ..., y^{(k-1)}] ,$$

$$S = [s^{(k-m_1)}, s^{(k-m_1+1)}, s^{(k-m_1+2)}, ..., s^{(k-1)}],$$

$$D = diag(s^{(k-m_1)} \cdot y^{(k-m_1)}, s^{(k-m_1+1)} \cdot y^{(k-m_1+1)}, s^{(k-m_1+2)} \cdot y^{(k-m_1+2)}, ..., s^{(k-1)} \cdot y^{(k-1)}),$$

$$(L)_{ij} = \begin{cases} s^{(k-m_1-1+i)} \cdot y^{(k-m_1-1+j)} & if\ i > j \\ 0 & otherwise \end{cases};$$

$\hat{B}^{(k)} \leftarrow B^{(k)}$ (project to free-variable sub space);

5) Determine the search direction:

$\hat{d}^{(k)} = -\hat{B}^{(k)} \backslash \hat{g}^{(k)}$, $d^{(k)} \leftarrow \hat{d}^{(k)}$ (project to the original space by filling zeros);

6) $x^{(k+1)} = x^{(k)} + \alpha d^{(k)}$. Normally, α is determined by a line search method; here, α can be a fixed value, e.g., α =1;

7) Check the lower limits of $x^{(k+1)}$, if not satisfied, set the corresponding components to be the lower limit;

8) Evaluate $u(x)$ at $x^{(k+1)}$;

9) If $m_1 < m$, then $m_1 = m_1 + 1$.

$s^{(k)} = x^{(k+1)} - x^{(k)}$, $y^{(k)} = g^{(k+1)} - g^{(k)}$, store $s^{(k)}$ and $y^{(k)}$.

$k = k + 1$.

Go to step 3);

In the above algorithm, steps 1~2 are to initialize the algorithm, steps 3~9 form the iteration to update the decision variables till reaching the optimum point. Steps 4~5 calculate the reduced Hessian matrix $\hat{B}^{(k)}$ and the direction $\hat{d}^{(k)}$. It is also possible to calculate the inverse Hessian matrix directly and using matrix multiplying to determine $\hat{d}^{(k)}$ [21].

VI. Case Study

A. Case study parameters

The case study is based on the Western System Coordinating Council (WSCC) 9-bus test case as shown in Fig. 2, which represents a simple approximation of the WSCC to an equivalent system with nine buses and three generators. One of the generators is replaced with a wind farm. The generator and line parameters are listed in Table I and Table II, respectively. The power transfer distribution factor (PTDF) can be calculated based on the line parameters.

The key parameters of the simulation are listed in Table III. The day-ahead ED problem is performed for 24 hourly based planning periods of the next day. The profiles of the total load and forecasted wind power are shown in Fig. 3. The real wind power is a random variable with the VPD distribution. The specific VPD parameters are related to the forecasted wind power and can be obtained from Table IV after standardizing the wind power forecast.

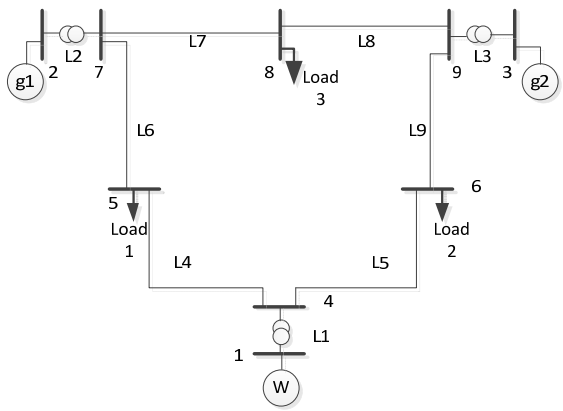

Fig. 2. Single line diagram of the WSCC 9-bus test case

TABLE I
GENERATOR DATA

a	b	c	Min (MW)	Max (MW)	Connect to bus	Max RD	Max RU
0.021	36.33	1658.57	0	300	2	80	80
0.018	38.27	1356.66	0	300	3	80	80

TABLE II
LINE PARAMETERS

left bus	right bus	X	Limit (MW)
4	1	0.0576	300
2	7	0.0625	300
9	3	0.0586	300
5	4	0.068	300
6	4	0.092	300
7	5	0.161	50
9	6	0.1738	300
7	8	0.0576	300
8	9	0.1008	300

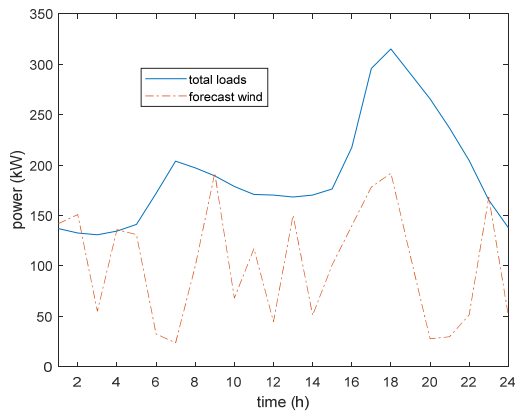

Fig. 3. The load and wind forecast profile

TABLE III
KEY PARAMETERS OF THE SIMULATION

parameter	value
Total required down reserve(MW)	80
Total required up reserve(MW)	80
Load 1 peak (MW)	125
Load 2 peak (MW)	90
Load 3 peak (MW)	100
Rated wind power, Max (MW)	200
k^{ov} , k^{un}	120, 60
c^{+} , c^{-} (confidence level)	0.95
ε (penalty)	0.005
Step size α (L-FBGS-B)	1
Step size α (subgradient method)	0.005
θ , m (L-FBGS)	1000, 5

TABLE IV
VPD DISTRIBUTION FOR DIFFERENT EXPECTED VALUE RANGE

When forecast in range	α	β	γ
[0.00, 0.04]	72.56	2.74	0.01
[0.04, 0.08]	66.22	2.14	0.04
[0.08, 0.12]	54.24	1.63	0.07
[0.12, 0.16]	50.54	1.57	0.12
[0.16, 0.20]	45.21	1.47	0.15
[0.20, 0.24]	38.84	1.37	0.19
[0.24, 0.28]	37.23	1.24	0.23
[0.28, 0.32]	36.14	1.15	0.28
[0.32, 0.36]	34.21	1.34	0.32
[0.36, 0.40]	33.47	1.29	0.36
[0.40, 0.44]	32.23	1.19	0.41
[0.44, 0.48]	31.89	1.13	0.45
[0.48, 0.52]	31.52	1.21	0.48
[0.52, 0.56]	32.14	1.01	0.52
[0.56, 0.60]	34.51	0.82	0.57
[0.60, 0.64]	36.74	0.79	0.62
[0.64, 0.68]	37.24	0.81	0.66
[0.68, 0.72]	37.51	0.77	0.69
[0.72, 0.76]	39.24	0.71	0.73
[0.76, 0.80]	42.21	0.67	0.78
[0.80, 0.84]	39.54	0.57	0.81
[0.84, 0.88]	65.15	0.38	0.88
[0.88, 0.92]	75.15	0.29	0.92
[0.92, 0.96]	88.53	0.21	0.95
[0.96, 1.00]	98.51	0.18	0.99

B. *Case study results*

The simulation was carried out using Matlab scripts on a PC with 8G RAM and 3.4GHz CPU. The subproblems are solved by the solver "fmincon" in Matlab with the interior point algorithm, while the master problem (L-BFGS-B) is solved by user programmed scripts.

1) Stochastic ED Results:

The stochastic ED results are shown in Fig. 4, Fig. 5 and Fig. 6. The total generator output (Fig. 4) matches the total load (Fig. 3). From the generator parameters, one can see that Generator g1 is cheaper than Generator g2; therefore, in most hours, g1 outputs more than g2. However, due to L6 congestion at hour 7, 20 and 21 (Fig. 5), g2 has to output more because g1 is contributing more to the congestion of L6 and has to reduce its output. Because the ramping limit, g2 has to output sufficient power before and after the congestion hours.

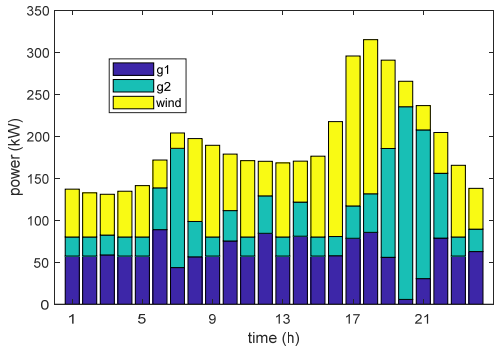

Fig. 4. The scheduled generator outputs

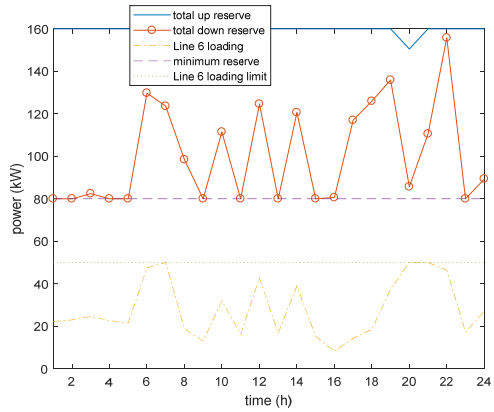

Fig. 5. The scheduled total up/down reserve and line loading of L6

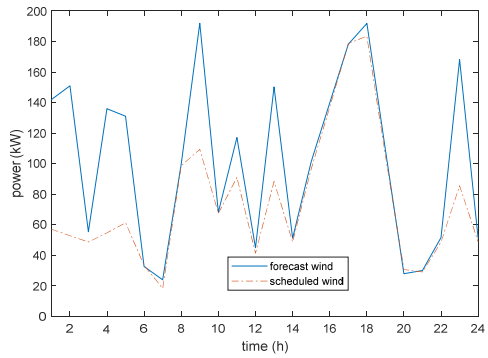

Fig. 6. The forecast wind and scheduled wind power

It can be seen from Fig. 5 that the total up/down reserve requirement is satisfied for all hours. The up/down reserve is calculated from the generator ramping limit and the difference between the generator output power and its max/min limit, whichever smaller. It should be noted that the auxiliary variables $r_{i,t}^{u}$ and $r_{i,t}^{d}$ are different (smaller) than the actual up/down reserve.

Compared to the forecasted wind power, the scheduled wind power is a bit smaller because the cost for overestimate k^{ov} (need up regulation from the regulation market) is bigger than the one for underestimate k^{un} (need down regulation). Such scheduling leads to a bigger chance of underestimate rather than overestimate.

2) Convergence Observation:

The key variables, including $\lambda_t, v_t^-, \omega_{k,t}^+$, their gradients and the generator output, are observed in the iteration process of the L-BFGS-B method. The results are shown in Fig. 8. It can be seen that all of them are settled down after about 150 iterations. Because the major time-consuming step is the evaluation of the objective function of the master problem, i.e., solving the sub-problems, the total time is determined by the number of the iterations. When settled down (reach the optimum), $\lambda_t, v_t^-, \omega_{k,t}^+$ should reach the values that satisfy the KKT optimal conditions. And the projected gradients, or the mismatches of the constraints ($df/d\lambda$ and df/dv in Fig. 8) should be close to zero, and the load flow should be within the line loading limit (equivalent to the mismatches being close to zero).

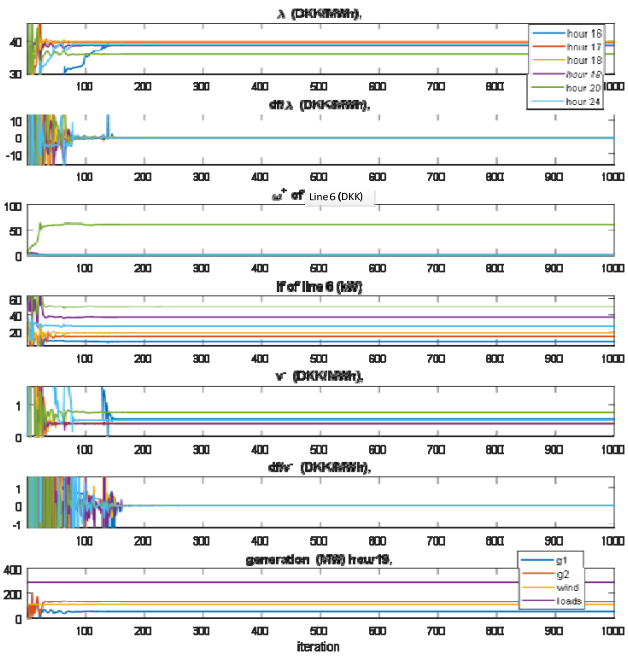

Fig. 7. Iteration observation of the key variables of L-BFGS-B method

Compared to the L-BFGS-B method, the normal subgradient method is quite slow. The convergence process is also observed, and the results are shown in Fig. 8. After 5000 iterations with the step size 0.005 (bigger step size is tested and failed to converge), some variables are close to but not yet reach the optimal point within the given tolerance. When close to the optimum, the mismatches will be small, and the production of the mismatches and the step size will be even smaller, leading to very small progress of each iteration in the last stage of the optimization process. This is why the subgradient method is slow. It can be seen that the final values of the variables are very close to those resulting from the L-BFGS-B method, which mutually verifies that both methods are reaching /approaching the correct optimum point.

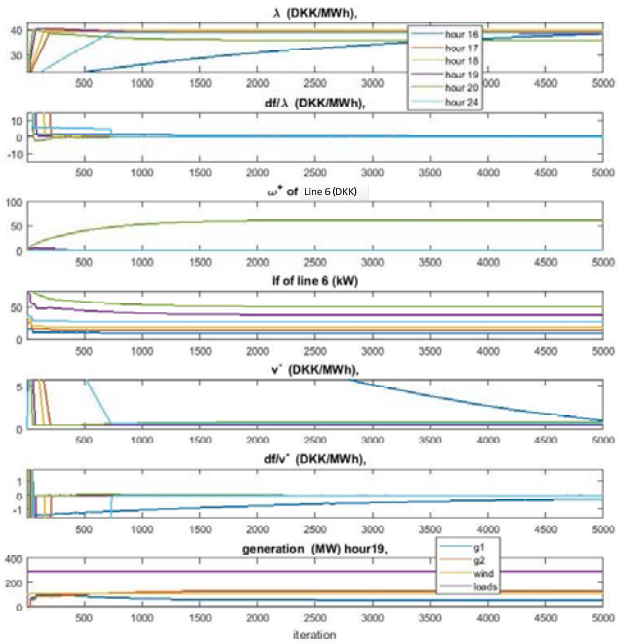

Fig. 8 Iteration observation of the key variables of normal subgradient method

C. Test on Nordic 32-bus system

The test on Nordic 32-bus system is mainly to show the scalability of the L-BFGS-B based dual decomposition method for solving stochastic ED problems with wind. The network structure is shown in Fig. 9, which consists of 4 zones, 32 buses connected by 52 lines and 8 transformers. There are 20 conventional generators. More details of the test system can be found in [26]. In addition, there are two cases considered in this study, one is with 5 wind farms and the other is with 20 wind farms randomly located in the 4 zones of the test system. Each wind farm has 100 MW and the total maximum load of the whole system is 10,000 MW.

The step size of the L-BFGS-B is set to 0.1. The results are shown in Table V. Two scenarios are considered for each case: one is with one transmission line (400kV) congestion, while the other is with congestion on multiple lines. The other limits are the same, such as the generation limits, ramping limits, balance and reserve requirement. The violation level in Table V refers to the percentage of the line flow that exceeds the limit. For instance, for the 800 MW limit, 10% violation means the line flow exceeds the limit by 80 MW. It can indicate how close the point is to the optimal solution. Because the dual problem is always a lower boundary of the primary

problem, the less the violation, the closer the point is to the optimal solution of the primary problem. It can be seen that, in general, more congestion lines mean more iterations needed for solving the problem. However, more wind farms do not mean more iterations due to the decomposition. In fact, in this study, it reduced the iteration number, but it is not general.

Compared to the 9-bus case, the iteration number doesn't change significantly. This means that the scalability of the L-BFGS-B method is good. On average, one iteration takes CPU time around 2 seconds covering the master problem and 20 (conv. generators) + 20 (wind farms) subproblems in parallel. However, if parallel techniques are not applied, one iteration will take 77~80 seconds for 40 subproblems; therefore, parallel techniques are recommended, especially for cases with large amounts of generators and wind farms.

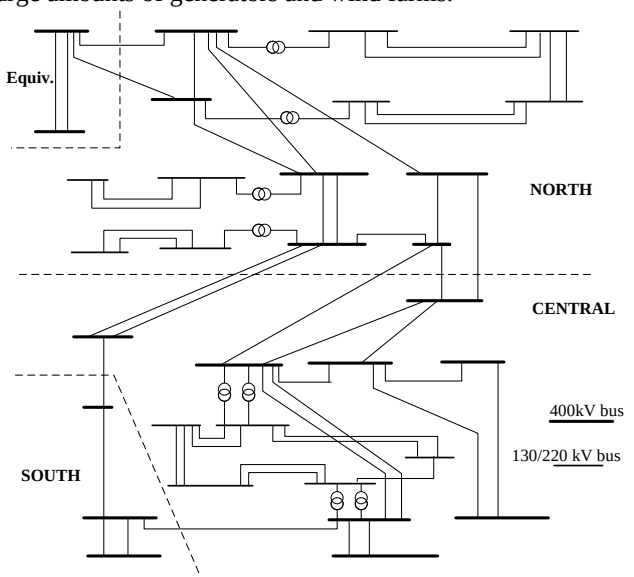

Fig. 9. Nordic 32-bus test system

TABLE V
ITERATIONS NEEDED TO REACH THE CORRESPONDING OPTIMUM LEVEL

viola-tion level	5 Wind Farms		20 Wind Farms	
	One con-gestion	multiple conges-tion	One con-gestion	multiple conges-tion
$\leq 10\%$	48	125	51	64
$\leq 1\%$	139	200	70	88
$\leq 0.1\%$	150	269	80	93

VII. CONCLUSIONS

This paper proved that the stochastic ED problem based on the VPD for integrating wind power is strictly convex. Because of the structure of the formulated stochastic ED problem, it is then decomposed into many sub-problems by the dual decomposition method, each of which handles one generator or wind farm. The dual decomposition makes the solution time not dependent on the problem scale. The decomposed problem can be solved by the subgradient method, which has a simple algorithm, or by the L-BFGS-B method, which has more efficient computation and faster iteration process. The case studies have validated the efficacy of the stochastic ED problem for scheduling generators and random wind power, and the efficiency of the dual decomposition and L-BFGS-B method.

REFERENCES

[1] B. H. Chowdhury and S. Rahman,"A review of recent advances in economic dispatch," *IEEE Trans. Power Syst.*, vol.5, no.4, pp. 1248–1259, 1990.

[2] F. N. Al Farsi, M. H. Albadi, N. Hosseinzadeh, and A. H. Al Badi,"Economic dispatch in power systems," in *Proc. 2015 IEEE 8th GCC Conference & Exhibition*, pp. 1–6.

[3] Z. Li, W. Wu, B. Zhang, H. Sun, and Q. Guo,"Dynamic economic dispatch using Lagrangian relaxation with multiplier updates based on a quasi-Newton method," *IEEE Trans. Power Syst.*, vol.28, no.4, pp. 4516–4527, Nov. 2013.

[4] Y. Wang, Q. Xia, and C. Kang,"Unit commitment with volatile node injections by using interval optimization," *IEEE Trans. Power Syst.*, vol.26, no.3, pp. 1705–1713, Aug. 2011.

[5] J. Hetzer, D. C. Yu, and K. Bhattarai,"An economic dispatch model incorporating wind power," *IEEE Trans. Energy Convers.*, vol.23, no.2, pp. 603–611, Jun. 2008.

[6] M. H. Albadi and E. F. El-Saadany,"Comparative study on impacts of wind profiles on thermal units scheduling costs," *IET Renew. Power Gener.*, vol.5, no.1, pp. 26-35, 2011.

[7] B.-M. Hodge and M. Milligan,"Wind power forecasting error distributions over multiple timescales," in *Proc. 2011 IEEE Power and Energy Society General Meeting*, pp. 1–8.

[8] B. C. Pal and R. A. Jabr,"Intermittent wind generation in optimal power flow dispatching," *IET Gener. Transm. Distrib.*, vol.3, no.1, pp. 66–74, Jan. 2009.

[9] V. S. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski,"A stochastic model for the optimal operation of a wind-thermal power system," *IEEE Trans. Power Syst.*, vol.24, no.2, pp. 940–950, May 2009.

[10] J. Wang, M. Shahidehpour, and Z. Li,"Security-constrained unit commitment with volatile wind power generation," in *Proc. 2009 IEEE Power & Energy Society General Meeting*, pp. 1-8.

[11] Z.-S. Zhang, Y.-Z. Sun, J. Lin, L. Cheng, and G.-J. Li,"Versatile distribution of wind power output for a given forecast value," in *Proc. 2012 IEEE Power and Energy Society General Meeting*, pp. 1–7.

[12] Z.-S. Zhang, Y.-Z. Sun, D. W. Gao, J. Lin, and L. Cheng,"A versatile probability distribution model for wind power forecast errors and its application in economic dispatch," *IEEE Trans. Power Syst.*, vol.28, no.3, pp. 3114–3125, Aug. 2013.

[13] L. Wang and C. Singh,"Balancing risk and cost in fuzzy economic dispatch including wind power penetration based on particle swarm optimization," *Electr. Power Syst. Res.*, vol.78, no.8, pp. 1361–1368, Aug. 2008.

[14] L. Wang and C. Singh,"Stochastic economic emission load dispatch through a modified particle swarm optimization algorithm," *Electr. Power Syst. Res.*, vol.78, no.8, pp. 1466–1476, Aug. 2008.

[15] L. Wang and C. Singh,"Stochastic combined heat and power dispatch based on multi-objective particle swarm optimization," in *Proc. 2006 IEEE Power Engineering Society General Meeting*, pp. 226-234.

[16] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley,"Notes on decomposition methods," 2008. Available: https://pdfs.semanticscholar.org/d746/f6b5f69401b147996d21a4b5940f6dd40020.pdf. [Accessed: 26-Mar-2017].

[17] S. Boyd, L. Xiao, and A. Mutapcic,"Subgradient methods," 2003. Available: https://web.stanford.edu/class/ee392o/subgrad_method.pdf. [Accessed: 27-Jun-2017].

[18] R. (Roger) Fletcher,*Practical methods of optimization*. Wiley, 1987.

[19] R. Fletcher and M. J. D. Powell,"A rapidly convergent descent method for minimization," *Comput. J.*, vol.6, no.2, pp. 163–168, Aug. 1963.

[20] C. G. BROYDEN,"The convergence of a class of double-rank minimization algorithms," *IMA J. Appl. Math.*, vol.6, no.3, pp. 222–231, Sep. 1970.

[21] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu,"A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol.16, no.5, pp. 1190–1208, Sep. 1995.

[22] M. Lange,"On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *J. Sol. Energy Eng.*, vol.127, no.2, p. 177-184, May 2005.

[23] F. Capitanescu, M. Glavic, D. Ernst, and L. Wehenkel,"Interior-point based algorithms for the solution of optimal power flow problems," *Electr. Power Syst. Res.*, vol.77, no.5–6, pp. 508–517, Apr. 2007.

[24] R. A. Jabr,"A primal-dual interior-point method to solve the optimal power flow dispatching problem," *Optim. Eng.*, vol.4, no.4, pp. 309–336, Dec. 2003.

[25] CVX Research Inc.. (2012, Aug.). CVX: Matlab Software for Disciplined Convex Programming, version 2.0. [online]. Available: http://cvxr.com/cvx/.

[26] PES-TR19. (2015). Test Systems for Voltage Stability Analysis and Security Assessment. [online]. Available: http://sites.ieee.org/pes-resource-center/files/2015/08/PES_TR19_Test-Systems-for-Voltage-Stability-Analysis-and-Security-Assessment1.pdf. [Accessed: 29-Nov-2017].